

(11)特許出願公開番号

特開2003-5774

(P2003-5774A)

(43)公開日 平成15年1月8日(2003.1.8)

(51)IntCl. ⁷	識別記号	F I	テーマコード(参考)
G 1 0 L 13/06		G 1 0 L 5/04	F 5 D 0 4 5
13/00		3/00	E
13/08			H

審査請求 未請求 請求項の数6 OL (全 8 頁)

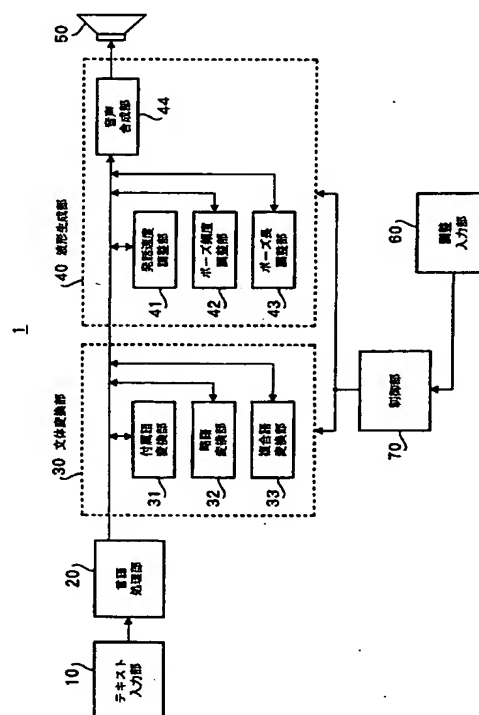
(21)出願番号	特願2001-191713(P2001-191713)	(71)出願人	000005821 松下電器産業株式会社 大阪府門真市大字門真1006番地
(22)出願日	平成13年6月25日(2001.6.25)	(72)発明者	加藤 弓子 大阪府門真市大字門真1006番地 松下電器 産業株式会社内
		(72)発明者	山上 勝義 大阪府門真市大字門真1006番地 松下電器 産業株式会社内
		(74)代理人	100105050 弁理士 鷲田 公一 Fターム(参考) 5D045 AA01

(54) 【発明の名称】 音声合成装置

(57) 【要約】

【課題】 入力されたテキストを音声に変換した際に、音声速度を標準の速度に対して大きく変化させた場合であっても、音声の自然性を損なわず、かつ、入力されたテキストの意味が正確にわかるようにすること。

【解決手段】 本発明は、使用者が入力した音声の速さが所定の範囲外にある場合に、音声に変換しようとするテキストをそのテキストと同等の意味を持ちかつ異なる表現を持つテキストに変換し、変換したテキストを音声に変換し出力し音声の単位時間あたりの情報量を調節することで音声の速度を調節するようにしたものである。



【特許請求の範囲】

【請求項1】 使用者が音声に変換するためのテキストを入力する入力手段と、使用者が音声の速さを段階的あるいは連続的に調節する速さ調節手段と、使用者が入力した音声の速さが所定の範囲外にある場合に、前記テキストを同等の意味を持ちかつ異なる表現を持つテキストに変換することで音声の単位時間あたりの情報量を調節する文体変換手段と、テキスト情報を音声に変換して出力する音声生成手段と、を具備したことを特徴とする音声合成装置。

【請求項2】 使用者が入力した音声の速さが所定の範囲内にある場合には、前記使用者が前記速さ調節手段より入力した音声の速さに応じて、出力する音声の単位時間あたりの音韻数を増減する発話速度調整手段、出力する音声に含まれるポーズの時間長を伸縮するポーズ長調整手段、および出力する音声に含まれるポーズの頻度を増減するポーズ頻度調整手段のいずれか1つの手段を用いて音声の速さを調節し、使用者が入力した音声の速さが所定の範囲外にある場合には前記発話速度調整手段、前記ポーズ長調整手段、および前記ポーズ頻度調整手段のいずれか1つの手段を用いて音声の速さを調節すると共に前記文体変換手段を用いて音声の速さを調整することを特徴とする請求項1に記載の音声合成装置。

【請求項3】 前記文体変換手段は言語情報に基づき冗長語の追加または削除または変更を行うものであることを特徴とする請求項1または請求項2の音声合成装置。

【請求項4】 前記文体変換手段は言語情報に基づき略語を省略されていない表現に変更するまたは特定の単語あるいは単語列を略語に変更するものであることを特徴とする請求項1または請求項2の音声合成装置。

【請求項5】 前記文体変換手段は言語情報に基づき付属語を含む特定の単語列を付属語を含まない複合語に変更するまたは特定の複合語を付属語を含む単語列に変換するものであることを特徴とする請求項1または請求項2の音声合成装置。

【請求項6】 指定された音声の速さが所定の範囲外にある場合に、音声に変換するためのテキストを同等の意味を持ちかつ異なる表現を持つテキストに変換することで音声の単位時間あたりの情報量を調節する工程と、前記異なる表現を持つテキストを音声に変換して出力する工程と、を具備したことを特徴とする音声合成方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、テキストを音声に変換する音声合成装置に関するものである。

【0002】

【従来の技術】従来より、テキストを音声に変換する音声合成装置がある。従来の音声合成装置は、音韻である音素の時間長を線形に伸縮することによって、音声速度を調節している。また、特開平06-266391号公

報には、母音重心点間の時間長のバランスを保持してそれぞれの音素時間長を伸縮することによって、音声速度を調節する方法が開示されている。これらの音声速度を調節する方法は、単位時間あたりの音韻数のみを調整することによって音声速度を調節している。

【0003】また、特許第3124791号公報には、発話速度に応じてポーズの頻度やポーズの時間長を調整することによって、音声速度を調節する方法が開示されている。

10 【0004】また、特開平9-311775号公報には、発話速度を変えるのではなく、使用者のテキストに対する参照レベル、つまり参照時間か参照速度によって決定されるユーザーの文書を参照する態度の指標に応じて、タイトルのみあるいは文書全体の音声に変換するテキストを削除することによって音声の速さを調節する方法が開示されている。

【0005】

【発明が解決しようとする課題】しかしながら、特開平06-266391号公報や特許第3124791号公報に記載された音声速度調整方法は、単位時間あたりの音韻数を変化させているが、発話全体の音韻の数を変化させない。すなわち、音韻あたりの時間長を変化させている。この音韻あたりの時間長がある所定の範囲を超えると、音韻を判別することが不可能になる。よって、このように単位時間あたりの音韻数を変化させる音声速度調整方法は、音声の自然性を損なわずに話速を変化させることのできる範囲が狭い。言い換えれば、音声速度を標準の速度に対して大きく変化させる場合には、音声の自然性が損なわれる可能性がある。

30 【0006】また、特開平9-311775号公報に記載された音声速度調整方法は、話速を上げるために、使用者のテキスト参照レベルに応じて一部を削除したテキスト部分のみを読み上げることになる。つまり、この方法は、入力されたテキストの情報すべてを音声に変換することができず、一部が削除されたテキストに含まれる情報の部分しか音声に変換できない。このため、使用者のテキスト参照レベルが標準に比べて大きく異なる場合は、削除するテキスト部分が多くなり、入力されたテキストの意味を正確に認識できない可能性がある。

40 【0007】本発明はかかる点に鑑みてなされたものであり、入力されたテキストを変換した音声を生成する際に、音声速度を標準の速度に対して大きく変化させた場合であっても、音声の自然性を損なわず、かつ、入力されたテキストの意味が正確にわかるようにすることを目的とする。

【0008】

【課題を解決するための手段】本発明は、使用者が入力した音声の速さが所定の範囲外にある場合に、音声に変換しようとするテキストをそのテキストと同等の意味を持ちかつ異なる表現を持つテキストに変換し、変換した

テキストを音声に変換し出力することで、音声の単位時間あたりの情報量を調節するようにしたものである。

【0009】このように、テキストを単純に削除するのではなく同等の意味を持つ異なる表現に変えたテキストを音声に変換することで、話速を変化させることなく、かつ入力されたテキストの意味が正確にわかるようにすることができる。

【0010】

【発明の実施の形態】本発明の第1の態様にかかる音声合成装置は、使用者が音声に変換するためのテキストを入力する入力手段と、使用者が音声の速さを段階的あるいは連続的に調節する速さ調節手段と、使用者が入力した音声の速さが所定の範囲外にある場合に、前記テキストを同等の意味を持ちかつ異なる表現を持つテキストに変換することで音声の単位時間あたりの情報量を調節する文体変換手段と、テキスト情報を音声に変換して出力する音声生成手段と、を具備した形態を採る。

【0011】この構成により、要約のように情報量を減少させることなく、口調としての音声の速度を自由に調整することが可能になる。この結果、入力されたテキストを変換した音声を生成する際に、音声速度を標準の速度に対して大きく変化させた場合であっても、音声の自然性を損なわず、かつ、入力されたテキストの意味が正確にわかるようにすることができる。

【0012】本発明の第2の態様は、第1の態様にかかる音声合成装置において、使用者が入力した音声の速さが所定の範囲内にある場合には、前記使用者が前記速さ調節手段より入力した音声の速さに応じて、出力する音声の単位時間あたりの音韻数を増減する発話速度調整手段、出力する音声に含まれるポーズの時間長を伸縮するポーズ長調整手段、および出力する音声に含まれるポーズの頻度を増減するポーズ頻度調整手段のいずれか1つの手段を用いて音声の速さを調節し、使用者が入力した音声の速さが所定の範囲外にある場合には前記発話速度調整手段、前記ポーズ長調整手段、および前記ポーズ頻度調整手段のいずれかひとつの手段を用いて音声の速さを調節すると共に前記文体変換手段を用いて音声の速さを調整するものである。

【0013】この構成により、使用者が入力した音声の速さが所定の範囲外にある場合には前記発話速度調整手段、前記ポーズ長調整手段、および前記ポーズ頻度調整手段のいずれかひとつの手段を用いて音声の速さを調節すると共に前記文体変換手段を用いて音声の速さを調整できる。これにより、情報量を減少させることなく、かつ発話速度だけでなく、口調としての音声の速度を自由に調整することが可能になる。

【0014】本発明の第3の態様は、第1の態様または第2の態様にかかる音声合成装置において、前記文体変換手段は言語情報に基づき冗長語の追加または削除または変更を行うものである。

【0015】この構成により、音声の自然性を損なわず、かつ、入力されたテキストの意味が正確にわかるようにできる。

【0016】本発明の第4の態様は、第1の態様または第2の態様にかかる音声合成装置において、前記文体変換手段は言語情報に基づき略語を省略されていない表現に変更するまたは特定の単語あるいは単語列を略語に変更するものである。

【0017】この構成により、音声の自然性を損なわず、かつ、入力されたテキストの意味が正確にわかるようにできる。

【0018】本発明の第5の態様は、第1の態様または第2の態様にかかる音声合成装置において、前記文体変換手段は言語情報に基づき付属語を含む特定の単語列を付属語を含まない複合語に変更するまたは特定の複合語を付属語を含む単語列に変換するものである。

【0019】この構成により、音声の自然性を損なわず、かつ、入力されたテキストの意味が正確にわかるようにできる。

【0020】本発明の第6の態様は、指定された音声の速さが所定の範囲外にある場合に、音声に変換するためのテキストを同等の意味を持ちかつ異なる表現を持つテキストに変換することで音声の単位時間あたりの情報量を調節する工程と、前記異なる表現を持つテキストを音声に変換して出力する工程と、を具備したことを特徴とする音声合成方法である。

【0021】以下、本発明の実施の一形態にかかる音声合成装置について説明する。まず、音声合成装置の構成について図1を用いて説明する。図1は、上記実施の形態にかかる音声合成装置の機能ブロック図である。

【0022】本実施の形態にかかる音声合成装置1は、音声に変換するテキストが入力されるテキスト入力部10が設けられている。テキスト入力部10は、入力されたテキストを出力する。言語処理部20は、テキスト入力部10より出力されたテキストを言語処理解析する。具体的には、言語処理部20は、テキストを形態素解析および構文解析し、少なくとも、読みおよびアクセントが含まれる音声合成用の中間言語および形態素解析結果、構文解析結果を出力する。

【0023】文体変換部30は、言語処理部20より出力された言語処理解析結果に基づき、テキスト入力部10で入力されたテキストと同等の意味を持ちかつ異なる表現を持つようにテキストを変換する。

【0024】文体変換部30は、言語情報に基づき冗長語の追加または削除または変更を行う付属語変換部31と、言語情報に基づき略語を省略されていない表現に変更するまたは特定の単語あるいは単語列を略語に変更する略語変換部32と、言語情報に基づき付属語を含む特定の単語列を付属語を含まない複合語に変更するまたは特定の複合語を付属語を含む単語列に変換する複合語変

換部33と、から構成される。

【0025】波形生成部40は、言語処理部20で解析されたテキスト、あるいは言語処理部20で解析され文体変換部30で表現を変換されたテキストの解析結果に基づき音声波形を生成する。

【0026】波形生成部40は、出力する音声の単位時間あたりの音韻数を増減する発話速度調整部41と、出力する音声に含まれるポーズの頻度を増減するポーズ頻度調整部42と、ポーズの時間長を伸縮するポーズ長調整部43と、ポーズ長調整部43の出力に応じて音声波形を生成する音声合成部44と、から構成される。

【0027】電気音響変換機50は、波形生成部40が生成した音声波形を音声に変換して出力する。調整入力部60は、使用者が出力音声の話速を調整するための信号を入力する部分である。

【0028】また、制御部70は、調整入力部60より入力された調整信号に従って文体変換部30、波形生成部40を制御する。具体的には、制御部70は、調整入力部60から入力された話速パラメータが標準話速に対して何倍かを判断する。そして、制御部70は、その倍率に応じて、発話速度調整部41、ポーズ頻度調整部42、およびポーズ長調整部43からひとつの処理部を選択し、選択した処理部を用いて音声波形を生成する。また、制御部70は、倍率が所定の範囲を超える場合は、付属語変換部31、略語変換部32、および複合語変換部33からひとつの処理部を選択し、選択した処理部を用いて言語処理解析したテキストを変換する。

【0029】以下、上記実施の形態にかかる音声合成装置の動作について図2を用いて詳細に説明する。図2は、上記実施の形態における音声合成装置の動作を示す流れ図である。

【0030】まず、制御部70が、テキストの入力に先立って調整入力部60より入力された話速パラメータを取得する(ステップ110)。次に、テキスト入力部10が、音声合成の対象となるテキストを受け付け、言語処理部20に出力する(ステップ120)。そして、言語処理部20は、テキスト入力部10より入力されたテキストを形態素解析および構文解析し、少なくとも、読みおよびアクセントが含まれる音声合成用の中間言語および形態素解析結果、構文解析結果を出力する(ステップ130)。

【0031】次に、制御部70は、ステップ110で取得した話速パラメータがあらかじめ定められた標準話速と等しいか判断する(ステップ140)。制御部70が話速パラメータと標準話速が等しいと判断した場合は、音声合成部44があらかじめ定められた標準発話速度、標準ポーズ頻度、標準ポーズ長を用い、ステップ130で言語処理部20より出力された中間言語に従って音声波形を生成し、出力する(ステップ150)。そして、電気音響変換機50が、音声合成部44が生成した音声

波形を音声に変換して出力する(ステップ160)。

【0032】一方、ステップ140で話速パラメータが標準話速と異なる場合、制御部70は、話速パラメータが標準の1/2、2倍から2.2倍の範囲内にあるか判断する(ステップ170)。ステップ170で話速パラメータが標準の1/2、2倍から2.2倍の範囲内にあると判断した場合は、発話速度調整部41が話速パラメータに応じて単位時間あたりの音韻数を調整した音声波形を生成する(ステップ180)。そして、電気音響変換機50が、音声合成部44が生成した音声波形を音声に変換して出力する(ステップ160)。

【0033】一方、ステップ170で話速パラメータが標準の1/2、2倍以下あるいは2.2倍以上であると判断した場合には、発話速度調整部41が単位時間あたりの音韻数に標準の1/2、2倍あるいは2.2倍に対応する音韻数を適用する(ステップ190)。しかしながら、話速パラメータが標準の1/2、2倍以下あるいは2.2倍以上である場合は、音韻数を単位時間あたりの音韻数に標準の1/2、2倍あるいは2.2倍に対応する音韻数を適用するだけでは、話速を十分に調節できていない。よって、音声合成装置1は、さらに、話速を調節する動作に移行する。

【0034】まず、制御部70は、話速パラメータと標準話速の1/2倍および2倍を比較する(ステップ200)。ステップ200において話速パラメータが標準話速の1/2倍から2倍の範囲内にあると判断した場合、ステップ190で調整された発話速度に対し、ポーズ頻度部42を用いてポーズ頻度を調整する処理、またはポーズ長調整部43を用いてポーズ長を調節する処理の少なくとも一方の処理を行い、生成した音声波形を出力する(ステップ210)。そして、電気音響変換機50が、音声合成部44が生成した音声波形を音声に変換して出力する(ステップ160)。

【0035】一方、ステップ200において話速パラメータが標準話速の1/2倍から2倍の範囲外にある場合、ポーズ頻度調整部42およびポーズ長調整部43は標準話速の1/2倍あるいは2倍に対応するポーズ頻度およびポーズ長を適用する(ステップ220)。しかしながら、話速パラメータが標準の1/2倍以下あるいは2倍以上である場合は、ポーズ頻度およびポーズ長に標準の1/2倍あるいは2倍に対応するポーズ頻度およびポーズ長を適用するだけでは、話速を十分に調節できていない。よって、音声合成装置1は、さらに、話速を調節する動作に移行する。

【0036】まず、制御部70は、話速パラメータが標準話速の1/2、5倍から2.5倍の範囲内にあるかを判断する(ステップ230)。ステップ230において話速パラメータが標準話速の1/2、5倍から2.5倍の範囲内にあると判断した場合、音声合成装置1は、ステップ130で言語処理部20より出力された中間言語

を、ステップ190で調整された発話速度と、ステップ220で調整されたポーズ頻度およびポーズ長を用いて処理した音声波形を音声出力する(ステップ160)。

【0037】一方、ステップ230において話速パラメータが標準話速の1/2.5倍から2.5倍の範囲外にあると判断した場合、制御部70は、さらに話速パラメータが標準の1/3.2倍から3.2倍の範囲内にあるか判断する(ステップ240)。そして、話速パラメータが標準の1/3.2倍から3.2倍の範囲内にある場合には、付属語変換部31が話速パラメータに応じて付属語の冗長度を調整する(ステップ250)。具体的には、付属語変換部31は、ステップ130で言語処理部20より出力された中間言語のうち付属語部分の中間言語を、調整された話速パラメータに対応する冗長度の付属語の情報に変換する。例えば、付属語変換部31は、話速パラメータが標準話速に比べて速い場合は、中間言語「です」を「だ」に変換する。一方、付属語変換部31は、話速パラメータが標準話速に比べて遅い場合は、中間言語「です」を「である」に変換するなどの変換をする。

【0038】一方、ステップ240において、話速パラメータが標準の1/3.2倍以下あるいは3.2倍以上の場合には、付属語変換部31は、標準話速の1/3.2倍あるいは3.2倍に対応する冗長度の付属語を、ステップ130で言語処理部20より出力された中間言語のうち付属語部分の中間言語に適用する。これにより、付属語変換部31は、付属語部分の中間言語を入力された話速パラメータに対応する冗長度の付属語の情報に変換する(ステップ260)。しかしながら、話速パラメータが標準の1/3.2倍以下あるいは3.2倍以上である場合は、標準の1/3.2倍あるいは3.2倍に対応する冗長度の付属語を、ステップ130で言語処理部20より出力された中間言語のうち付属語部分の中間言語に適用するだけでは、話速を十分に調節できていない。よって、音声合成装置1は、さらに、話速を調節する動作に移行する。

【0039】まず、制御部70は、話速パラメータが標準話速の1/3倍から3倍の範囲内にあるかを比較する(ステップ270)。ステップ270において話速パラメータが標準話速の1/3倍から3倍の範囲内にあると判断した場合、ステップ190で調整された発話速度と、ステップ220で調整されたポーズ頻度およびポーズ長と、ステップ260で変換された中間言語に従って音声波形を生成し、生成した音声波形を音声にして出力する(ステップ160)。

【0040】一方、ステップ270において話速パラメータが標準話速の1/3倍から3倍の範囲外にあると判断した場合、制御部70は、話速パラメータが標準の1/3.3倍以上であるか3倍以上であるか判断する(ステップ280)。

【0041】ステップ280において、話速パラメータが標準の1/3.3倍以上であるか3倍以上であると判断した場合には、略語変換部32が話速パラメータに応じて文中の略語の頻度を調整する(ステップ290)。具体的には、略語変換部32は、ステップ130で言語処理部20より出力された中間言語が調整された話速パラメータに応じた文中の略語の頻度になるように、略語に変換可能な複合語に対応する中間言語を略語に変換、あるいは、略語に対応する中間言語の部分を複合語に変換する(ステップ290)。例えば、略語変換部32は、音声の出力時間を短くする場合には、「第1研究開発部」という複合語を「第1研」などのような略語に変換し、音声の出力時間を長くする場合には、「第1研」という略語を「第1研究開発部」などのような複合語に変換する。

【0042】そして、変換した中間言語を音声波形にし、音声として出力する(ステップ160)。

【0043】さらに、制御部70は、話速パラメータと標準話速の1/3.3倍とを比較する(ステップ300)。ステップ300において話速パラメータが標準話速の1/3.3倍以上であると判断した場合、ステップ190で調整された発話速度と、ステップ220で調整されたポーズ頻度およびポーズ長と、ステップ260、ステップ290で変換された中間言語に従って音声波形を生成し、生成した音声波形を音声にして出力する(ステップ160)。

【0044】一方、ステップ220において、話速パラメータが標準話速の1/3.3倍を下回ると判断した場合、複合語変換部33が話速パラメータに応じて文中の複合語の品度を調整して、ステップ130で言語処理部20より出力された中間言語のうち複合語に対応する中間言語の部分に対して、複合語を構成する単語の切れ目にあたる位置に付属語の情報を挿入する(ステップ310)。例えば、複合語変換部33は、音声の出力時間を長くする場合は「参加登録書類」という複合語を「参加の登録の書類」というように付属語を挿入し、音声の出力時間を短くする場合は「参加の登録の書類」という付属語で接続された表現を「参加登録書類」というように付属語を削除する。

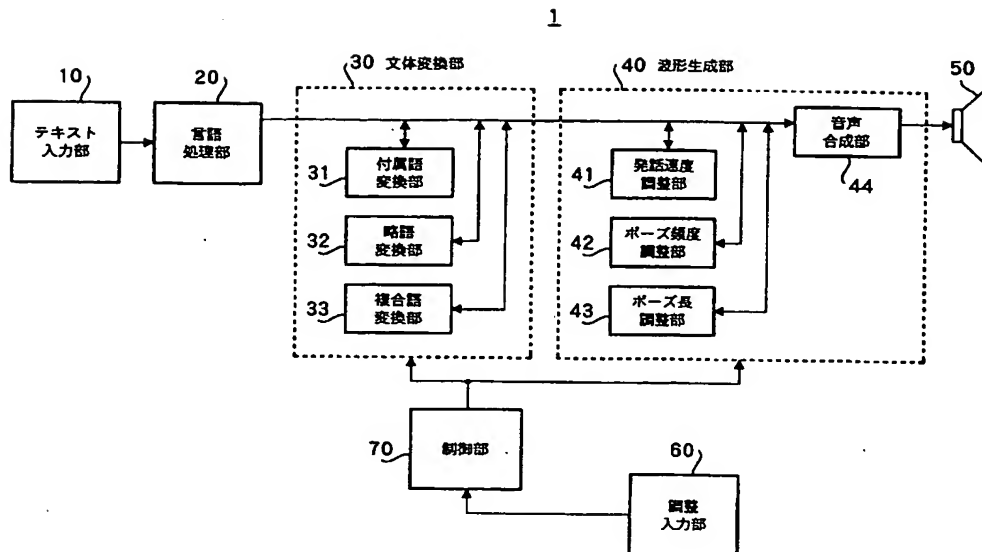
【0045】そして、音声合成装置1は、ステップ190で調整された発話速度と、ステップ220で調整されたポーズ頻度およびポーズ長と、ステップ260、ステップ290およびステップ310で変換された中間言語に従って音声波形を生成し、生成した音声波形を音声にして出力する(ステップ160)。

【0046】次に、上記実施の形態の話速パラメータ設定方法および話速調節方法について図3を用いて説明する。図3は、話速調整の入力インタフェースの1例と入力に対応する話速調整を実現するための方式を示したものである。

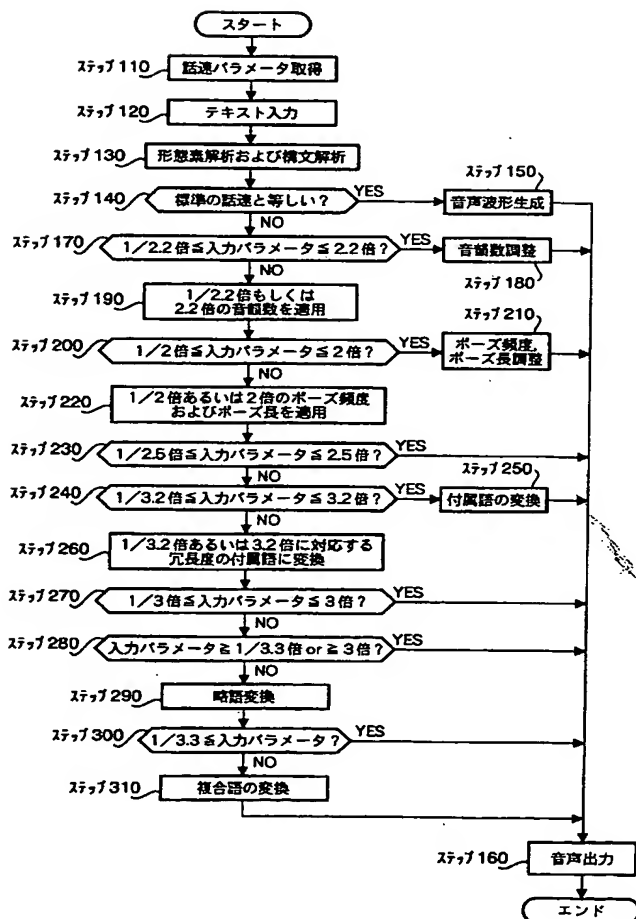
【0050】以上説明したように、上記実施の形態によれば、要約のように情報量を減少させることなく、かつ発話速度だけでなく、口調としての音声の速度を自由に調整することが可能になる。この結果、入力されたテキストを変換した音声を生成する際に、音声速度を標準の

70 制御部

【図1】



【図2】



【図3】

